Supplementary Information for "Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control project consortium."

Supplementary Notes

Supplementary Note 1: Previous studies

For amplicon sequencing, few studies have investigated the end-to-end effects of multiple protocol steps[1-4], instead focusing on one or a few features. As a rough approximation, and representing by no means an exhaustive review, these can be classified into sample collection and storage[5-7]; nucleotide extraction, amplicon primer choice, and PCR amplification[8-11]; sequencing strategy and platform[12-14]; contamination[15, 16]; and bioinformatics[17-20]. These reports make it clear that each step of microbial community amplicon sequence analysis can produce changes in taxonomic profiles. However, no previous study has quantified the extent of these effects among different protocols at multiple centers, nor contextualized under what circumstances each might dominate.

Supplementary Note 2: Bioinformatic protocols

MBQC-base OTU calling strategies consisted of three main approaches. Classification, perhaps most common for closed- or mixed-reference calling in QIIME[38], refers to joint clustering of sample and reference 16S rRNA gene sequences, with taxonomy assignment based on co-clustering between them. Clustering, more common for open-reference calling (e.g. as in UPARSE[30]), clusters only sample sequences, performing taxonomic assignment using post-hoc lookup of one representative sequence per OTU cluster. Mapping, finally, refers to direct lookup of all individual sequences against a reference database. For MBQC-base, primary taxonomic assignments were carried out using any of these algorithms against Greengenes 13.5, but several labs chose to include sequences not classified during this process in additional open-reference OTUs using arbitrary identifiers and an extended phylogeny.

Supplementary Note 3: Taxonomic composition

Taxonomic profiles of the MBQC-base samples were comparable to those of published Western gut microbiomes[1, 2], with the exception of the oral artificial communities, as expected (**Supplementary Fig. 1**). 10 phyla and 63 genera were detected with a minimum relative abundance of 0.01% in at least 10% of samples, with the major gut bacterial phyla (Firmicutes and Bacteroidetes) accounting for an average of 81% of phyla. Proteobacteria were present at an average of 11%, while Actinobacteria, Verrucomicrobia, Fusobacteria, and Cyanobacteria accounted for the majority of remaining taxa (7% average relative abundance in total).

Supplementary Note 4: Extraction kit considerations

Experimental sample handling protocol variation had the greatest effect on alpha diversity, particularly extraction kit. However, it is important to note that because we did not specify protocols for individual laboratories to use, these results are confounded by kit choice. For example, the Omega BioTek extraction kit was only used in laboratory HL-D, and Promega kit was only used in laboratory HL-I, which meant that the contribution of kit choice to variation in outcomes could not be differentiated from other variables.

Supplementary Note 5: Quantifying extraction kit variation

To estimate the effect of DNA extraction kit, we investigated the 18 unique specimens for which at least one paired sequencing reaction was performed by the same handling laboratory for locally and centrally extracted DNA. We quantified the degree to which the abundance distribution of each phylum differed between locally- and centrally-extracted samples within each handling lab, considering only phyla with at least 1% average relative abundance (Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria and Verrucomicrobia). For each bioinformatics protocol, we performed a paired t-test between samples within handling lab in 14 distinct combinations of extracting and sequencing laboratory. In general, when the MO-BIO kit was used, p-values are much larger, indicating that when the same kit was used by two different extracting labs (including the DNA centrally extracted using the same kit), results tended to be more similar (**Supplementary Fig. 9**).

Supplementary Note 6: Negative controls

Evaluation criteria for the “accuracy” of negative controls are non-obvious, because sequencing provides only relative count measurements. A small number of contaminant cells from many different organisms will result in high diversity; high contaminant biomass from a single organism will yield low diversity, but is not necessarily “better.” Some library construction protocols will preferentially deplete sequences from low-biomass samples,

but not consistently. In fact, negative control samples tended to rank among the lowest depth samples from some handling labs (e.g. HL-F, HL-J, HL-B) but among the highest in others (e.g. HL-A, HL-K). Successful recent efforts to control microbiome sequencing contamination[27] have used negative controls to identify potential contaminants, whose relative abundances must subsequently be assessed in non-control samples to determine the directionality of transfer (biospecimen-to-control, environment-to-control, reagents-to-biospecimen, environment-to-biospecimen, etc.)

Supplementary Note 7: Population-scale microbiome studies

At population scales, the MBQC aims to support at least two tasks not yet easy to carry out : well-controlled design of large scale microbiome population studies (e.g. tens to hundreds of thousands of samples from multiple sites and sources), and quantitative meta-analyses of multiple microbiome datasets. With respect to experimental design, our results, in combination with previous studies, suggest several recommendations. All protocol variables (handling or bioinformatic) of potentially large effect must be matched across sites and kept constant during long-running population studies. Some protocol steps e.g. specific fixation method, storage, equipment manufacturer, or reagent lot may have small effects under most circumstances, but, in specific microbial communities, still contribute to substantial variation. Multiple positive and negative controls should be included with each sample batch and retained in computational analyses. Even near-identical protocols can produce divergent results when used in multiple centers, particularly with respect to environmental or within-batch contaminants. Finally, while commercial kits can be used for some steps, where they will improve consistency when used (e.g. sample collection and extraction), other steps have few standard options (e.g. primer and index sets and/or bioinformatics still have many unique in-house components).

Supplementary Note 8: Specific protocol variables of note

MBQC-base results called out two additional variables of importance: PCR protocol (not primer selection) and bioinformatics. In the MBQC-base, the effects of amplification-related variables other than primers were small, although detectable (e.g. greater misclassification rates attributable to indexing strategy, **Supplementary Fig. 13**). We were not powered to assess PCR enzymes or programs, as these tended to differ among all labs; using Illumina technologies, rates of chimerism were low in all cases (**Supplementary Fig. 13**). Studies in which these variables were specifically assessed tended to focus on only a single laboratory environment and on much larger differences in enzyme, cycle count, or mate pairing strategy than were included here[9, 21, 29], allowing overall smaller effect sizes to be detected as significant. Finally, differences in computational protocol were mixed across all studies, in agreement with our finding that bioinformatics comparability is crucially dependent on which data processing steps vary and how their effects are reported.

Supplementary Note 9: MBQC-II

Since distinct human body sites and populations can harbor extremely different microbial communities, the habitat sources (e.g. skin), physical sampling methods (e.g. swabs, commercial kits), and sample shipping and storage conditions (e.g. fixation, freezing) for epidemiological studies must be investigated. Long-read sequencing platforms such as PacBio and Oxford Nanopore must be tested, along with different experimental approaches such as shotgun metagenomics, metatranscriptomics or multi-omics (e.g. metabolomics, proteomics). Computationally, minimum reporting standards do not exist for reproducible microbiome bioinformatics workflows; these are challenging to develop, but are important.

Supplementary Tables

Strain #	HMP number (where relevant)	Species	Family{21497273}	Approx. loopfuls (in 500mL 0.9% saline)	CFU/500 mL*
ORAL Community					
BT 1B CT2	HM-233	*Bacillus licheniformis*	Bacillaceae	4	3.00E+11
12_1_47 B FAA	HM-30	*Bifidobacterium sp.*	Bifidobacteriaceae	6	4.50E+11
CC87LB		*Rothia mucilaginosa*	Brevibacteriaceae	6	4.50E+11
CD1 D6 FAA 3		*Parvimonas micra*	Clostridia Incertae Sedis XI	3	2.25E+11
CD1 D5 FAA 3		*Mogibacterium timidum*	Clostridia Incertae Sedis XIII	<0.1	<7500000000
10_1_50	HM-213	*Campylobacter concisus*	Campylobacteriaceae	3	2.25E+11
MR1 #12		*Eggerthella lenta*	Coriobacteriaceae	3.5	2.63E+11
CD1 D6 FAA 13		*Slackia exigua*		3.5	2.63E+11
1_1_55	HM-44	*Klebsiella pneumoniae*	Enterobacteriaceae	6	4.50E+11
1_A_54 (D10)	HM-41	*Fusobacterium periodonticum*	Fusobacteriaceae	6	4.50E+11
4_A_31 (D28)	HM-196	*Leptotrichia goodfellowii*	Leptotrichiaceae	5	3.75E+11
F16 #1		*Weissella cibaria*	Leuconostocaceae	5	3.75E+11
CC92I		*Eikenella corrodens*	Neisseriaceae	4	3.00E+11
GT 4A CT1	HM-234	*Neisseria mucosa*		6	4.50E+11
6_1_58 FAA CT1	HM-227	*Tannerella sp.*	Porphyromonadaceae	6	4.50E+11
CC98A		*Prevotella oralis*	Prevotellaceae	6	4.50E+11
CC21 001D		*Capnocytophaga sputigena*	Ravobacteriaceae	0.5	3.75E+10
CC94D		*Granulicatella adiacens*	Streptococcaceae	6	4.50E+11
2_1_36 FAA	HM-60	*Streptococcus gordonii*		6	4.50E+11
CC57F		*Gemella morbillorum*	Tenericutes Incertae Sedis XI	5	3.75E+11
CD1 D5 FAA 6		*Dialister pneumosintes*	Veillonellaceae	<0.1	<7500000000
3_1_44	HM-64	*Veillonella parvula*		5	3.75E+11
				95.5	
GUT Community					
29_1	HM-85	*Coprobacillus cateniformis*	Anaeroplasmataceae	3.5	2.63E+11
F16 #22		*Bifidobacterium angulatum*	Bifidobacteriaceae	6	4.50E+11
4_8_47 FAA	HM-304	*Collinsella aerofaciens*	Coriobacteriaceae	6	4.50E+11
AC2/8/11 AN D5 FAA 1		*Bilophila wadsworthia*	Desulfovibrionaceae	2.25	1.69E+11
1_1_43	HM-37	*Escherichia coli*	Enterobacteriaceae	6	4.50E+11
3_1_5R	HM-65	*Fusobacterium gonidiaformans*	Fusobacteriaceae	6	4.50E+11
12_1B	HM-57	*Fusobacterium varium*		6	4.50E+11
5_1_63 FAA	HM-152	*Anaerostipes hadrus*	Lachnospiraceae	6	4.50E+11
CC43 001B		*Clostridium bolteae*		6	4.50E+11
7_1_47 FAA	HM-228	*Lactobacillus iners*	Lactobacillaceae	3.5	2.63E+11
7_4A	HM-172	*Pediococcus acidilactici*		4	3.00E+11
5_7_47 FAA	HM-185	*Ralstonia pickettii*	Oxalobacteriaceae	6	4.50E+11
CC33 002B		*Paenibacillus barengoltzii*	Paenibacillaceae	3.5	2.63E+11

UC1 BHI R		*Parabacteroides merdae*	Porphyromonadaceae	6	4.50E+11
MR1 #13		*Bacteroides caccae*	Bacteroidaceae	6	4.50E+11
5_U_42 FAA	HM-184	*Propionibacterium acnes*	Propionibacteriaceae	6	4.50E+11
ETR2 #14		*Alistipes shahii*	Rikenellaceae	3.5	2.63E+11
6_1_47 FAA	HM-226	*Subdoligranulum variabile*	Ruminococcaceae	1.5	1.13E+11
30_1	HM-323	*Enterococcus saccharolyticus*	Streptococcaceae	6	4.50E+11
22-5-S 12 D6 FAA		*Synergistes sp.*	Synergistetes	2	1.50E+11

Supplementary Table 1: Gut- and oral-derived artificial community compositions. Bacterial strains used, along with their approximate relative abundances based on loop counts. Artificial communities were developed from in-house isolates representative of the range of bacterial genera found within the oral or gut habitats. All strains were originally isolated from human subjects and HMP genome reference strains were included where available[21]. *: Calculated by multiplying approx. loopfuls by 7.5e10.

Sample type	Health status	Sex	Age	BMI	# Fecal	# DNA	DNA Conc. (ng/ul)	Abs 260/280
Fresh	Sick (ICU)	Unknown	Unknown	Unknown	3	2	54	1.94
Fresh	Sick (Diabetes/RA)	Female	25	Unknown	2	0	NA	NA
Fresh	Healthy	Male	36	Unknown	2	3	49	1.80
Fresh	Healthy	Male	37	Unknown	2	2	61	1.94
Fresh	Sick (ICU)	Unknown	Unknown	Unknown	3	2	5.4	1.19
Fresh	Sick (ICU)	Unknown	Unknown	Unknown	2	2	4.7	1.47
Fresh	Sick (ICU)	Female	40	Unknown	2	0	NA	NA
Fresh	Sick (Diabetes/RA)	Female	25	Unknown	3	2	34	1.77
Fresh	Sick (ICU)	Unknown	Unknown	Unknown	2	2	15	1.72
Fresh	Healthy	Female	2	Unknown	2	2	27.6	1.65
Fresh	Healthy	Female	2	Unknown	2	0	NA	NA
Freeze-dried	6 month post-surgery CRC case	Male	70	25.06	3	2	48	1.90
Freeze-dried	6 month post-surgery CRC case	Male	65	24.33	2	3	64	1.74
Freeze-dried	4 month post-surgery CRC case	Male	64	32.28	2	0	NA	NA
Freeze-dried	Pre-surgery control	Male	50	33.45	3	2	101	1.86
Freeze-dried	6 month post-surgery CRC case	Male	67	22.08	2	2	27	1.73
Freeze-dried	Pre-surgery control	Male	30	21.59	2	3	115	1.82
Freeze-dried	3 month post-surgery CRC case	Male	56	22.71	2	2	79	1.89
Chemostat (time 1)	Healthy	Male	25	Unknown	3	2	24.7	2.60
Chemostat (time 2)	Healthy	Male	25	Unknown	3	2	14.4	2.90
Oral artificial	NA	NA	NA	NA	3	3	44.5	1.82
Fecal artificial	NA	NA	NA	NA	3	3	42.4	1.93
Blank	NA	NA	NA	NA	0	2	NA	NA

Supplementary Table 2: Aliquots derived from physical specimens used for handling and data generation in the MBQC-base. All laboratories in the handling module were provided with one or more copies of the same aliquot set, each uniquely blinded. The 96-aliquot set was derived from 22 physical specimens (11 fresh stool, seven freeze dried, two chemostat, and two artificial), a subset of which were centrally extracted in addition to being shipped to participants in raw form. Replication and triplication yielded 41 aliquots of extracted DNA, 53 aliquots of raw stool, and two negative control aliquots of storage buffer. For centrally extracted DNA, concentrations (ng/ul) and Abs 260/280 values were obtained using a NanoDrop ND-8000.

bioinformatics lab	median	mean	sd
BL-3	1341.00	1311.75	544.41
BL-9B	18049.00	30044.17	33601.19
BL-9A	26722.00	43136.13	45408.83
BL-2	23973.50	39842.69	42882.86
BL-1	26530.50	42675.93	45322.21
BL-7	24053.00	58199.82	114359.97
BL-4	23573.00	43852.68	60844.12
BL-8	25202.00	40484.67	40936.89
BL-5	27109.00	59848.61	120182.39
BL-6	25733.50	40794.42	54919.45

Supplementary Table 3: Per-bioinformatics lab summary of the number of reads assigned across all samples. As an indication of each bioinformatics lab's QC stringency, we show the mean, median and standard deviation of reads used per sample across all samples processed by each bioinformatics lab. Two outliers indicate the labs that included open reference OTUs instead of Greengenes OTU identifiers, thus preserving many more sequences in miscellaneous OTUs in some samples.

Specimen Type	Bioinformatics Lab	Inverse Simpson	Chao1	Observed Species
blank	BL-3	8.25 (3.97)	224.65 (135.22)	152.32 (87.54)
Human	BL-3	6.89 (3.14)	247.12 (128.58)	160.44 (73.95)
Artificial	BL-3	8.68 (2.4)	94.8 (69.26)	60.79 (33.96)
Chemostat	BL-3	6.53 (2.68)	184.59 (62.1)	124.97 (33.77)
blank	BL-2	7.79 (3.64)	1396.68 (1763.84)	853.78 (1050.89)
Human	BL-2	6.21 (3.26)	1992.11 (1894.76)	1285.47 (1163.07)
Artificial	BL-2	8.88 (3.01)	1217.35 (1272)	755.33 (676.64)
Chemostat	BL-2	6.94 (1.93)	1958.34 (1568.98)	1275.03 (921.15)
blank	BL-1	7.78 (3.96)	307.07 (168.04)	254.4 (151.04)
Human	BL-1	5.92 (3.07)	328.11 (147.86)	277.4 (131.62)
Artificial	BL-1	8.67 (2.8)	210.37 (127.32)	153.5 (97.84)
Chemostat	BL-1	6.86 (1.94)	316.98 (118.79)	266.39 (97.1)
blank	BL-4	8.02 (3.8)	185.78 (124.4)	160.41 (113.6)
Human	BL-4	6.32 (3.12)	211.03 (112.27)	180.63 (95.94)
Artificial	BL-4	8.33 (2.88)	125.07 (102.24)	92.91 (80.66)
Chemostat	BL-4	6.16 (2.63)	192.61 (90.79)	163.33 (73.54)
blank	BL-8	10.5 (4.91)	602.68 (445.54)	441.81 (327.81)
Human	BL-8	8.09 (4.07)	834.71 (534.68)	635.23 (387.09)
Artificial	BL-8	9.38 (2.72)	477.3 (424.83)	331.36 (264.82)
Chemostat	BL-8	7.01 (2.3)	780.14 (422.28)	595.37 (285.75)
blank	BL-6	8.8 (3.83)	604.56 (465.46)	445.43 (359.04)
Human	BL-6	7.1 (3.53)	811.26 (516.11)	615.56 (384.93)
Artificial	BL-6	8.95 (3.06)	457.46 (410.03)	299.43 (267.73)
Chemostat	BL-6	6.66 (2.9)	663.36 (399.18)	486.12 (264.56)
blank	BL-9B	7.91 (3.19)	155.11 (105.5)	154.69 (105.23)
Human	BL-9B	6.09 (3.19)	184.08 (96.48)	182.89 (95.72)
Artificial	BL-9B	8.89 (2.65)	78.79 (66.27)	78.58 (66.04)
Chemostat	BL-9B	6.17 (1.77)	154.48 (58.13)	153.81 (57.84)
blank	BL-9A	8.24 (4.04)	851.76 (825.33)	605.91 (604.13)
Human	BL-9A	6.1 (3.21)	1538.09 (1240.93)	1031.9 (817.17)
Artificial	BL-9A	8.73 (2.98)	867.98 (625.64)	539.8 (346.14)
Chemostat	BL-9A	5.43 (2)	1444.76 (865.04)	945.47 (501.47)

Supplementary Table 4: Per-bioinformatics lab summary of alpha diversity stratified by sample type. Diversity measures summarize all samples (aggregated across all handling labs) that passed QC in each bioinformatics lab. Parentheses indicate standard deviations.

Kruskal-Wallis p-values, beta-diversities between specimens, stratified by sample type (human, chemostat, or artificial - not handling/bioinformatics)	
	1.37E-66
Kruskal-Wallis p-values, beta-diversities between specimens, stratified by handling lab (not sample type/bioinformatics)	
	0
Kruskal-Wallis p-values, beta-diversities between specimens, stratified by bioinformatics lab (not sample type/handling)	
	0
Wilcoxon p-values, beta-diversities between specimens, each handling lab vs. all others (not stratified by sample type/bioinformatics)	
HL-K	2.32E-62
HL-C	6.39E-08
HL-I	0.0582786
HL-J	3.04E-50
HL-A	1.07E-199
HL-E	4.88E-33
HL-H	0.0289539
HL-L	5.00E-121
HL-N	2.69E-11
HL-B	2.67E-49
HL-F	4.89E-33
HL-D	3.38E-105
HL-M	0.837295
Wilcoxon p-values, beta-diversities between specimens, each bioinformatics lab vs. all others (not stratified by sample type/handling)	
BL-3	0.00367346
BL-9B	9.27E-14
BL-2	0.00764754
BL-1	0.0226126
BL-4	7.27E-42
BL-8	5.83E-215
BL-9A	0.372145
BL-6	0.0371751

Supplementary Table 5: Omnibus and pairwise univariate significance of specimen type, handling, and bioinformatics laboratories. All tests use between-specimen beta-diversities from **Fig. 3**. Tests are as stated in table, all p-values are nominal.

	InvSimpson	Chao1	Observed Species	PD
InvSimpson	1	0.322440375	0.372269434	0.405390207
Chao1	0.322440375	1	0.974760524	0.909345409
Observed Species	0.372269434	0.974760524	1	0.945236634
PD	0.405390207	0.909345409	0.945236634	1

Supplementary Table 6: Alpha diversity measures correlation. Spearman correlation matrix between $\log_{10}$ transformed alpha metrics for 20,708 samples (see **Supplementary Fig. 4**).

Variable	Description	Coding
Specimen number	The unique specimen aliquoted to each lab	Unique ID for each specimen
Handling laboratory	The lab which performed the sequencing of the sample	Unique ID for each lab
Bioinformatics laboratory	The lab which performed the bioinformatic analyses of the reads	Unique ID for each lab
Pre-extraction	Whether the aliquot was sent to the lab pre-extracted or not	Pre extracted vs. not
Specimen type	The format of the specimen	Freeze-dried vs. oral artificial colony vs. fecal artificial colony vs. robogut vs. fresh
Health status	Whether the samples came from healthy or sick individuals	Sick (ICU/Diabetes/RA/CRC case) vs. not sick (control, healthy, and mock communities)
16S primer	The 16S PCR primer set	EMP V4 515F/806R vs. Schloss 2013 vs. Other
Read length	Targeted read length when sequencing	Continuous values
Sequencing chemistry version	The version of the sequencing chemistry used	V2 vs. V3 vs. V1/V4
phiX fraction	Fraction of phiX included during sequencing	Continuous values
Fraction quality bases	Average fraction of bases of acceptable quality (per Illumina)	Continuous values
Log read count	Natural logarithm of the total sequenced read count	Continuous values
OTU software	Summary indicator for software used to call OTUs	QIIME (1.8, 1.9) vs. UPARSE 7 vs. other
OTU clustering	True if unsupervised clustering (rather than supervised mapping) was used for OTU calling	Yes vs. No
Taxonomic Assignment Method	The methodology used for assigning taxonomy to reads	Classification vs. mapping vs. clustering
OTU filtering	Whether the OTUs were filtered after calling	Yes vs. No

Supplementary Table 7: Variables used for multivariate model of variance partitioning in microbiome handling and bioinformatics protocols. Table includes the names of each variable included in the complete model, a brief description, and the levels or continuous range used for coding.

Sample ID	Library Input (ng)	Library Protocol
7165328	100	Kapa auto-6 PCR cycles
6033534	100	Kapa auto-6 PCR cycles
9886767	100	Kapa auto-6 PCR cycles
6893758	100	Kapa auto-6 PCR cycles
2790814	100	Kapa auto-6 PCR cycles
2209563	100	Kapa auto-6 PCR cycles
DZ153190158	17.6	Kapa - manual; 12 PCR cycles
DZ153190153	7.4	Kapa - manual; 12 PCR cycles
DZ153190152	12.3	Kapa - manual; 12 PCR cycles
DZ153190186	< 5 ng	Kapa - manual; 12 PCR cycles
DZ153190199	< 5 ng	Kapa - manual; 12 PCR cycles
DZ153190217	< 5 ng	Kapa - manual; 12 PCR cycles

Supplementary Table 8: Fecal and oral artificial community replicate sample handling for shotgun metagenomic validation. Replicates of the fecal and oral positive control specimens were centrally shotgun sequenced to quantify detectable organisms independently of amplification and 16S amplicon primer effects. Table lists the 12 replicates sent (six oral and six fecal, six pre-extracted and six raw), input DNA quantification, and subsequent library construction protocol.

Supplementary Datasets

Supplementary Dataset 1: Extended literature review and categorization of microbiome protocol studies. A list of all Pubmed-indexed studies since 2010 evaluating protocols for human and/or artificial microbial community 16S amplicon data analysis. For inclusion, publications were required to compare multiple protocols, target whole communities (not single microbial genomes), and to be neither phylogenetically-targeted (i.e. as bacterially universal as possible) nor specific to 454 pyrosequencing (although comparisons to such data were included). Studies are, as best as possible, categorized into reviews, integrative analyses of multiple protocol factors, or the approximate stages of: positive control evaluations in the form of one or more mock communities; sample collection and storage or fixation; nucleotide extraction; PCR primer or variable region selection; other aspects of PCR amplification (e.g. enzyme, cycles, index design); sequencing platform; negative controls or contaminant analysis; and computational protocols.

Supplementary Dataset 2: Handling protocol variables recorded during the MBQC-base. Raw responses to the MBQC-base sample handling protocol form. Major protocol variables captured included the lab identifier, DNA extraction kit, homogenization steps, 16S amplification primer set, reagents, sequencing platform and chemistry, PhiX percentage, total sequencing yield, and quality base rate.

Supplementary Dataset 3: Bioinformatics protocol variable reporting. All bioinformatics labs received re-blinded, demultiplexed FASTQ files from all sample handling labs, which they processed into OTU tables with standard identifiers (plus additional open reference OTUs if desired). Bioinformatics protocols were particularly diverse, and the main protocol variables reported here include per-read QC/filtering, mate pair stitching software and parameters, OTU construction steps, and per-OTU, per-sample, or whole-dataset QC/filtering.

Supplementary Dataset 4: Handling protocol variables curated from the MBQC-base. Curated responses from **Supplementary Dataset 2**, in which extraneous or poorly reported variables have been removed, units have been normalized, incorrect values removed or replaced, and formatting improved.

Supplementary Dataset 5: Raw read counts per sample. For each sample handling lab, the number of raw reads per sample at deposition (prior to any bioinformatics).

Supplementary Dataset 6: MBQC-base OTU table. Greengenes v13.5 taxon identifiers as provided by all MBQC bioinformatics labs for each sample passing all per-lab quality control steps. Per-sample metadata also appended include the originating physical specimen, sample type, handling lab, basic handling and bioinformatics protocol variables, bioinformatics lab, and ecological summary statistics (alpha-diversity etc.)

Supplementary Dataset 7: Alpha and beta-diversities stratified by sample type within and between handling and bioinformatics laboratories. Raw data corresponding to **Fig. 3**, in which Bray-Curtis dissimilarities within labs are computed only between technical replicates handled and extracted identically; between lab distributions compare only replicates from the same originating specimen as processed by one lab to all others. Beta diversities represent a random subsample of all pairwise comparisons for each category for reasons of efficiency in computing the overall distribution.

Supplementary Dataset 8: Beta-diversity divergence from artificial community positive controls, stratified by handling and bioinformatics laboratories. Raw data corresponding to **Fig. 4**, in which genus-level Bray-Curtis distances from the two reference positive control communities are summarized (20 fecal and 22 oral isolates, respectively), stratified between centrally and locally extracted samples and by sample handling and bioinformatics laboratory.

References

1. Brooks, J.P. et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* **15**, 66 (2015).
2. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. & Schloss, P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **79**, 5112-5120 (2013).
3. Jones, M.B. et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci U S A* **112**, 14024-14029 (2015).
4. Degnan, P.H. & Ochman, H. Illumina-based analysis of microbial community diversity. *The ISME journal* (2011).
5. Song, S.J. et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. *mSystems* **1** (2016).
6. Lauber, C.L., Zhou, N., Gordon, J.I., Knight, R. & Fierer, N. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol Lett* **307**, 80-86 (2010).
7. Hill, C.J. et al. Effect of room temperature transport vials on DNA quality and phylogenetic composition of faecal microbiota of elderly adults and infants. *Microbiome* **4**, 19 (2016).
8. Gohl, D.M. et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol* **34**, 942-949 (2016).
9. Walker, A.W. et al. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* **3**, 26 (2015).
10. Kennedy, N.A. et al. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* **9**, e88982 (2014).
11. Soergel, D.A., Dey, N., Knight, R. & Brenner, S.E. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* **6**, 1440-1444 (2012).
12. Clooney, A.G. et al. Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLoS One* **11**, e0148028 (2016).
13. Schloss, P.D., Jenior, M.L., Koumpouras, C.C., Westcott, S.L. & Highlander, S.K. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* **4**, e1869 (2016).
14. Benitez-Paez, A., Portune, K.J. & Sanz, Y. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION portable nanopore sequencer. *GigaScience* **5**, 4 (2016).
15. Salter, S.J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* **12**, 87 (2014).
16. Lauder, A.P. et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* **4**, 29 (2016).
17. Schloss, P.D., Gevers, D. & Westcott, S.L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* **6**, e27310 (2011).
18. Edgar, R.C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**, 996-998 (2013).
19. Kopylova, E. et al. Open-Source Sequence Clustering Methods Improve the State Of the Art. *mSystems* **1** (2016).
20. Schmidt, T.S., Matias Rodrigues, J.F. & von Mering, C. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ Microbiol* **17**, 1689-1706 (2015).
21. Nelson, K.E. et al. A catalog of reference genomes from the human microbiome. *Science* **328**, 994-999 (2010).